Apprentissage Automatique I MATH60629

Les fondements de l’apprentissage automatique — **Sommaire** — Semaine #2

Un problème d’apprentissage automatique

Les trois ingrédients d’un problème de ML :

1. Tâche. Quel est le problème à résoudre?

- Modèle. Comment paramétrer la solution.

2. Performance. Le modèle est-il bon?

3. Expérience. À quel type de données avez-vous accès?

Les types d’expériences

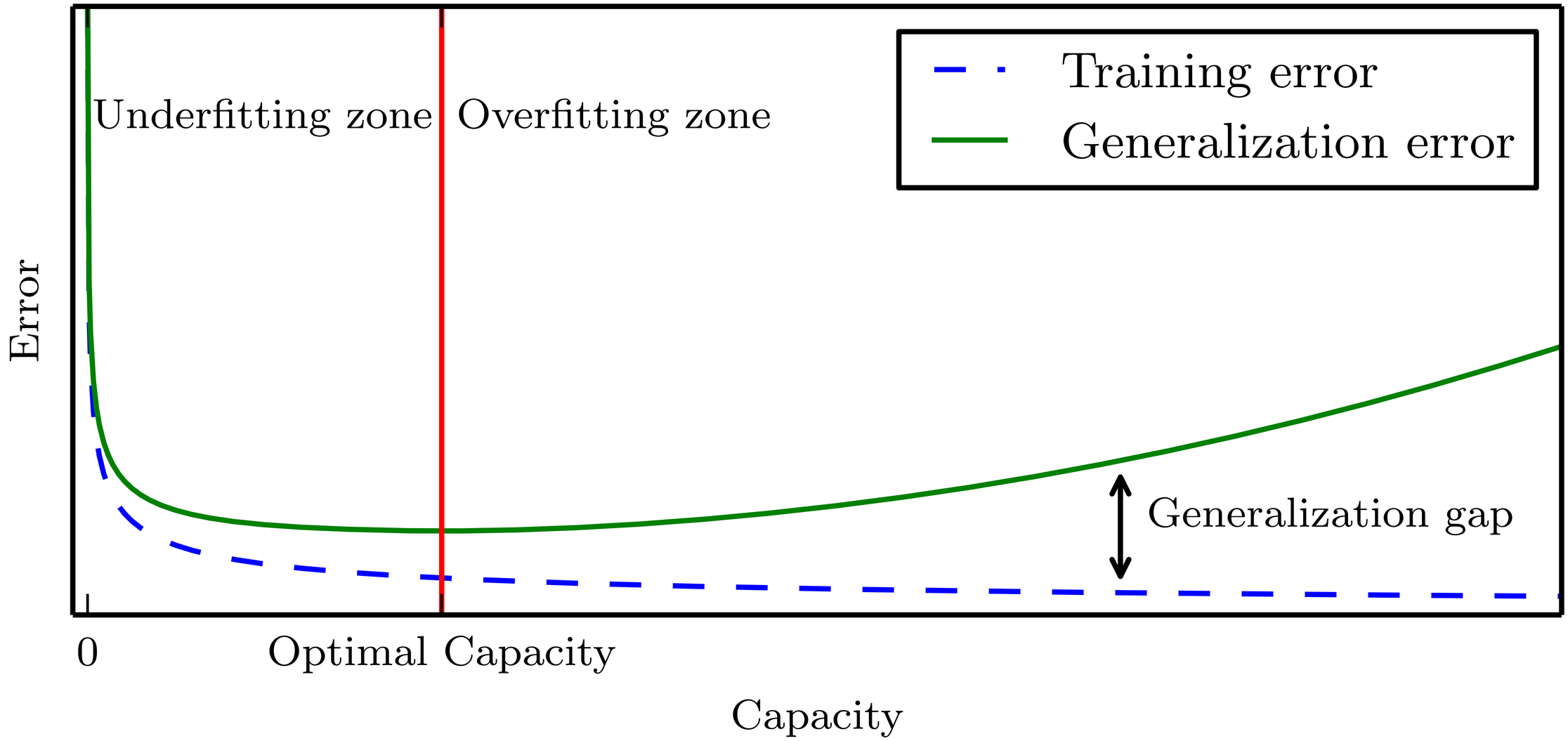
A. **Supervisé {(x,y)}.** P. ex., régression, classification. f: X -> Y

B. **Non supervisé {(x)}.** P. ex., regroupement, réduction de la dimension, estimation de la densité.

C. **Par renforcement.** Un agent pose des actions dans un environment.

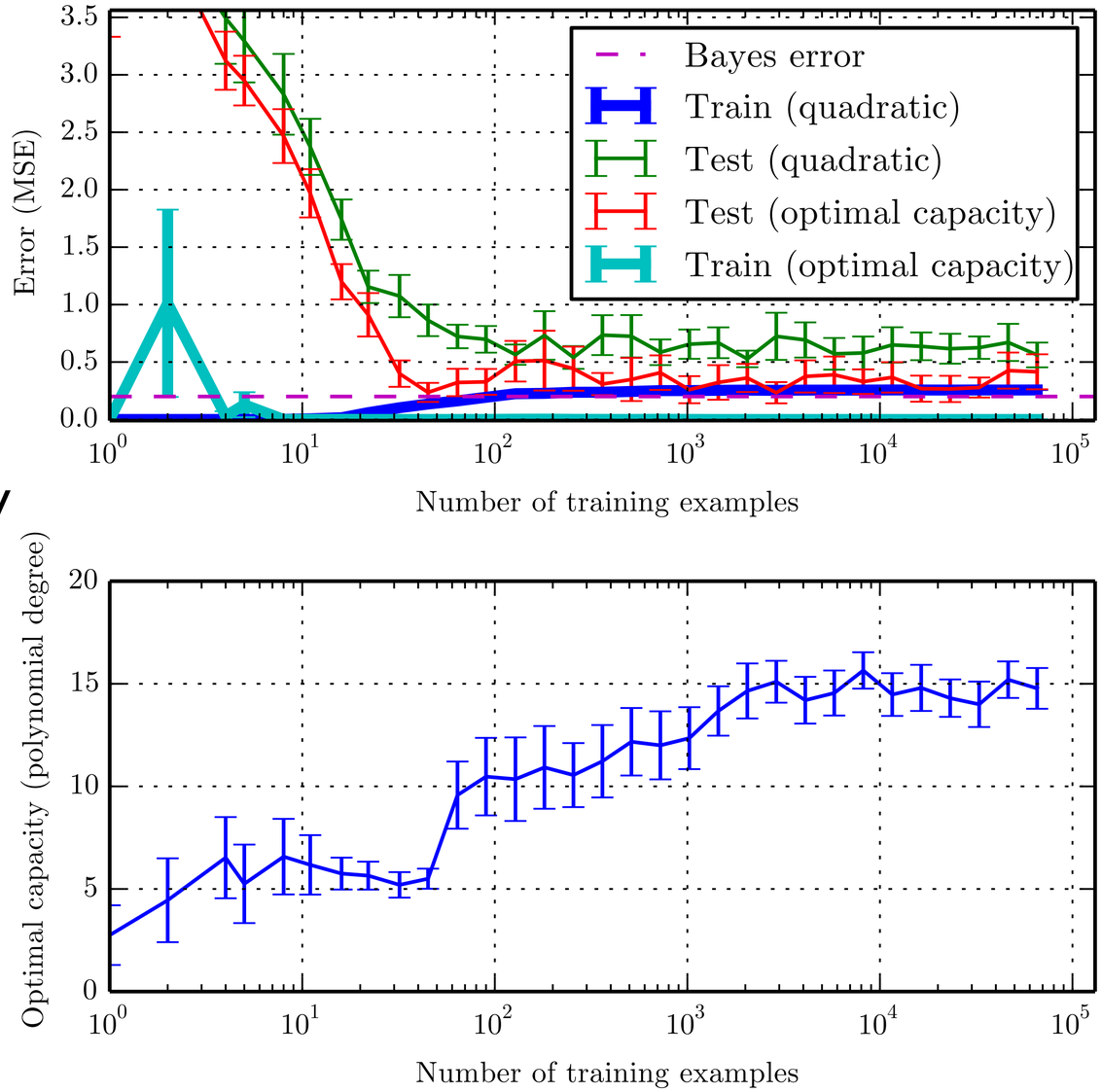
Évaluation d’un modèle

- On nous donne:
  - Une mesure de la performance
  - Un ensemble d’entraînement
  - Un modèle
- On peut calculer:
  - L’erreur d’entraînement: on l’utilise pour apprendre.
  - L’erreur d’entraînement **ne peut pas** être utilisée pour évaluer votre modèle
    - Vous devez utiliser un ensemble de données séparé (par exemple un ensemble de validation)


Underfitting zone
Overfitting zone
Training error
Generalization error
Error
Generalization gap
0
Optimal Capacity
Capacity

Synthetic data is generated using a degree 5 polynomial $y = w_5x^5 + w_4x^4 + w_3x^3 + w_2x^2 + w_1x^1$

Training set size also plays an important role in a model's capacity to generalize



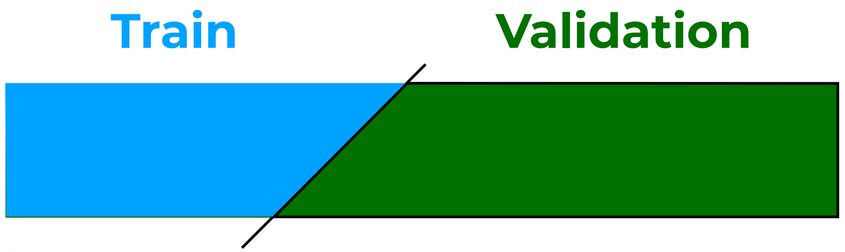
[Figure 5.4, Chapter 5, Deep Learning]

Regularisation

- C’est une façon de limiter la capacité d’un modèle
- Loss $:= \text{MSE}^{\text{train}} + \lambda \mathbf{w}^\top \mathbf{w}$

Ensemble de Validation

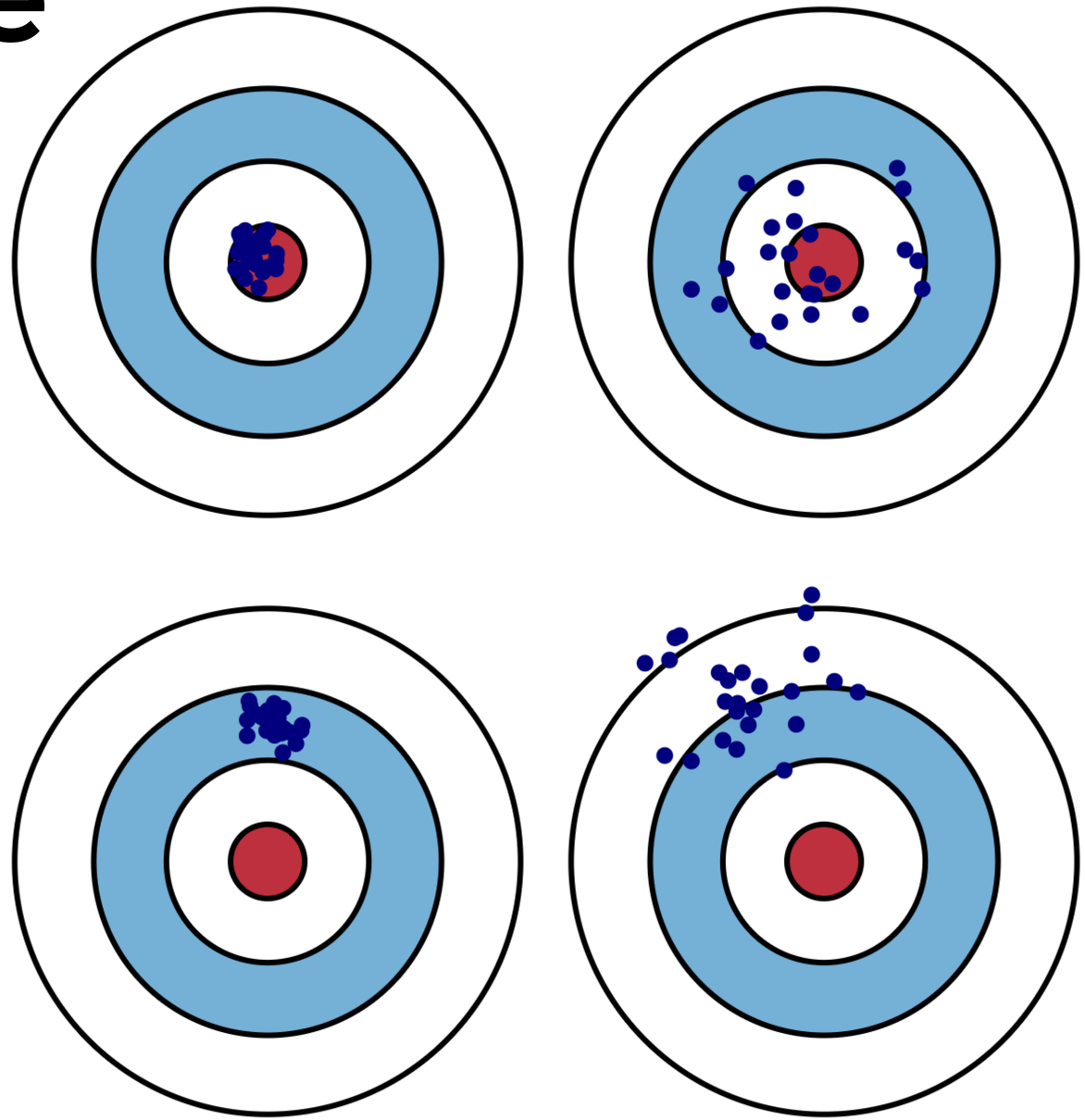
- Comment choisir le bon modèle ou les bons hyperparamètres (p. ex. λ)?
  - En utilisant un ensemble de validation
    - Diviser les données en deux:
      1. Ensemble d'entraînement
      2. Ensemble de validation
        - C'est un proxy pour l'ensemble de test

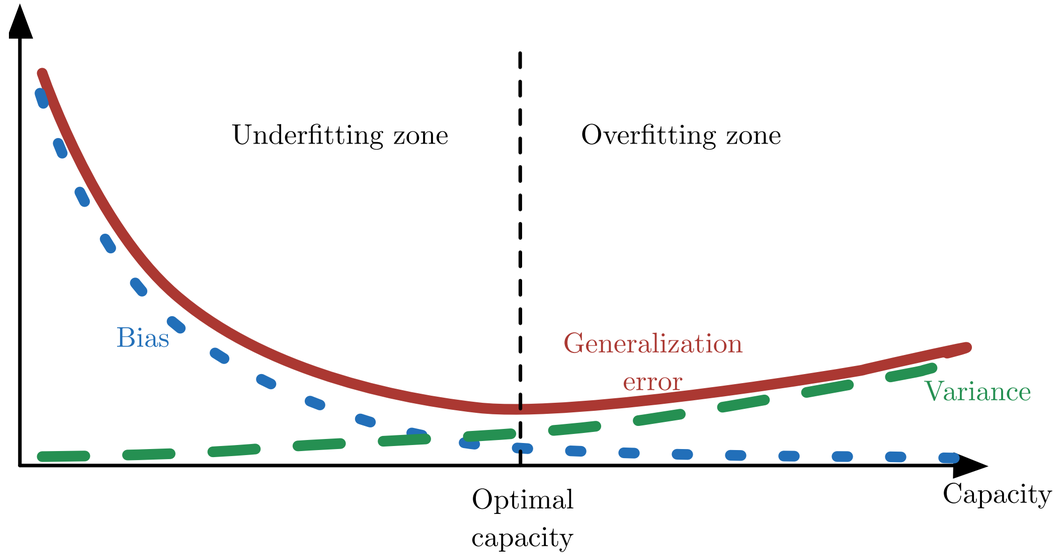


  - Entraînez différents modèles ou hyperparamètres sur l'ensemble d'entraînement
  - Choisissez le meilleur modèle selon sa performance sur l'ensemble de validation

Biais / Variance

- Le but est d'atteindre le centre de la cible (en rouge)
- Chaque point bleu représente la "performance" d'un modèle sur un ensemble de données venant d'une distribution fixe

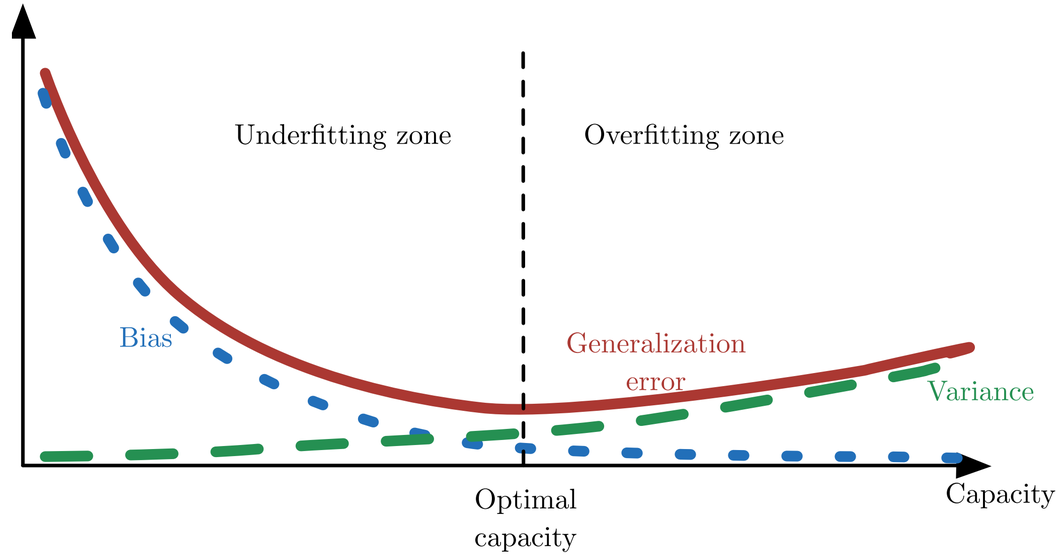


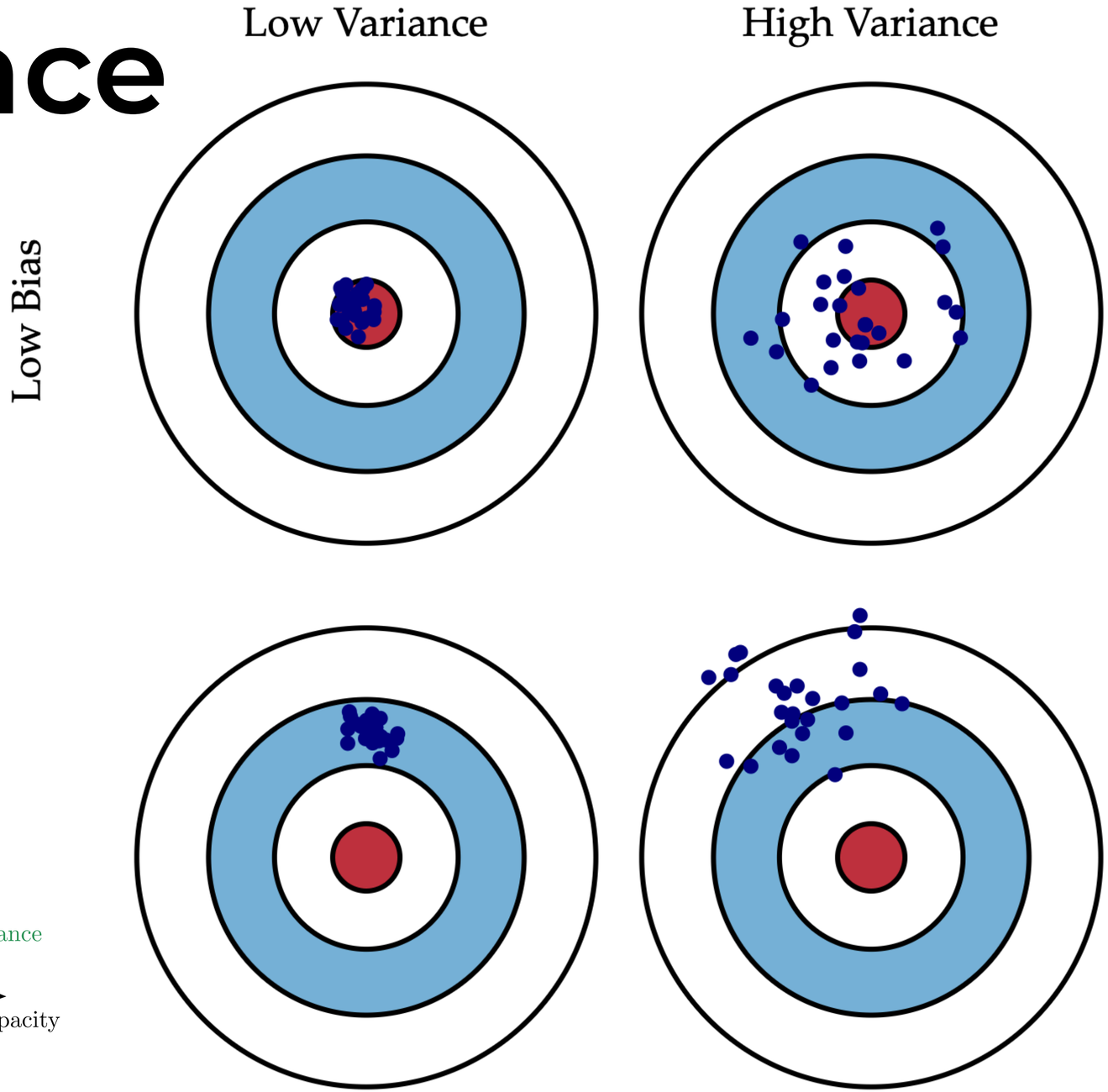


http://scott.fortmann-roe.com/docs/BiasVariance.html

Biais / Variance

- Le but est d'atteindre le centre de la cible (en rouge)
- Chaque point bleu représente la "performance" d'un modèle sur un ensemble de données venant d'une distribution fixe





http://scott.fortmann-roe.com/docs/BiasVariance.html